DOCUMENT RESUME

ED 426 099 TM 029 329

AUTHOR Amado, Alfred J.
TITLE Partitioning Predicted Variance into Constituent Parts: A Primer on Regression Commonality Analysis.
PUB DATE 1999-01-00
NOTE 18p.; Paper presented at the Annual Meeting of the Southwest Educational Research Association (San Antonio, TX, January 21-23, 1999).
PUB TYPE Reports - Descriptive (141) -- Speeches/Meeting Papers (150)
EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Heuristics; *Prediction; *Regression (Statistics)
IDENTIFIERS *Commonality Analysis; Partitioning Procedures; *Variance (Statistical)

ABSTRACT

Commonality analysis is a method of decomposing the R squared in a multiple regression analysis into the proportion of explained variance of the dependent variable associated with each independent variable uniquely and the proportion of explained variance associated with the common effects of one or more independent variables in various combinations. Unlike other variance partitioning methods (e.g., stepwise regression) that distort the results, commonality analysis considers all possible orders of entry into the model and does not depend on a priori knowledge to arrange the predictors. However, traditionally commonality analyses have been underutilized in research. The purpose of this paper is to introduce commonality analysis as an accurate and efficient method for partitioning variance. A data set is used to provide a heuristic example that explains the steps and guidelines necessary for performing a commonality analysis. Tables provide visual aids. (Contains 2 tables and 15 references.) (Author/SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 426 099

Partitioning Predicted Variance into Constituent Parts

A Primer on Regression Commonality Analysis

Alfred J Amado

Texas A&M University

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
This document has been reproduced as received from the person or organization originating it.
Minor changes have been made to improve reproduction quality.
Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Alfred Amado

TM029329

Paper presented at the annual meeting of the Southwest Educational Research Association, San Antonio, TX, January, 1999

Abstract

Commonality analysis is a method of decomposing the R^2 in a multiple regression analysis into the proportion of explained variance of the dependent variable associated with each independent variable uniquely and the proportion of explained variance associated with the common effects of one or more independent variables in various combinations. Unlike other variance partitioning methods (e.g., stepwise regression) that distort the results, commonality analysis considers all possible orders of entry into the model and does not depend on a priori knowledge to arrange the predictors. However, traditionally commonality analyses have been underutilized in research. The purpose of the present paper is to introduce commonality analysis as a accurate and efficient method for partitioning variance. A data set is used to provide a heuristic example that explains the steps and guidelines necessary for performing a commonality analysis. Tables are utilized to provide visual aids.

Partitioning Predicted Variance into Constituent Parts

A Primer on Regression Commonality Analysis

Initially, ANOVA was developed to relieve researchers of the computational burden inherent in analyzing data, because it partitions the variance of the dependent variable uncorrelated parts it provides computational simplicity (Cohen, 1968). In fact, all OVA methods (i.e., ANOVA, ANCOVA, MANOVA) convert intervally-scaled independent variables into nominally-scaled independent variables, even when these variables are not already nominally-scaled (Thompson, 1984). However, increased access to computers and rapid advancements in computer software have led to reduced dependence on OVA methods. As a result, researchers and analysts do not need to continue the prodigal discarding of variance that occurs when intervally-scaled variables are converted to nominally-scaled variables (Murthy, 1994), which consequently leads to a loss of information and a less sensitive analysis (Pedhazur, 1982). In addition, it has been shown that OVA-type methods also reduce the reliability of the variables considered in the design, inflate Type II error probability, and distort the distribution shapes and relationships among the variables (Cohen, 1968; Murthy, 1994; Thompson, 1992).

These shortcomings have resulted in less use of OVA methods by researchers (cf. Elmore & Woehlke, 1988; Goodwin & Goodwin, 1985; Willson, 1980) and a greater use of general linear model approaches such as regression (Rowell, 1991, 1996; Thompson, 1992). As Thompson (1992) explains, the increased usage of multiple regression is due, in part, to the realization that all parametric statistical analyses are part of a single general linear model (e.g., regression, canonical correlation analysis, and structural equation modeling). As Neter, Kutner, Nachtsheim,

and Wasserman (1996) noted, "Regression analysis is a statistical methodology that utilizes the relation between two or more quantitative variables so that one variable can be predicted from the other, or others" (p. 3). Regression analyses are of great use in identifying the unique contribution of each predictor variable in explaining the variance of the dependent variable. The interpretation of a regression analysis is fairly straightforward when the design consists of only one predictor variable, as in Case I, or when the predictor variables are perfectly uncorrelated, as in Case II. When there is no overlap (i.e., perfectly uncorrelated) between the predictor variables, the sum of the squared bivariate correlations (r^2) for the predictors is equal to the squared multiple correlation (R^2) involving all the predictors (Thompson, 1992).

Therefore, the partitioning of variance in Cases I and II is relatively easy and straightforward to interpret. However, models that have predictor variables that are correlated to some extent, which is usually the case, provide greater complications, making it more difficult to determine the "true" effects of the independent variables on the dependent variable. Of particular concern is the fact that the sum of the squared simple correlations rarely sums to the squared multiple correlation (Beaton, 1973). As Thompson and Borrello (1985) emphasized, in such instances it is necessary to examine both beta weights and structure coefficients when interpreting such data. However, examining the beta weights and the structure coefficients does not explain the relative contribution of each predictor, uniquely or in combination, with other predictors, in the regression analyses.

Researchers can better understand the contribution of each predictor variable with the use of methods that partition the variance of R^2 into all the constituent parts that can be attributed to

each predictor variable (Rowell, 1991, 1996). One method for conducting this partitioning of variance is by performing a commonality analysis on the data. Commonality analysis, also referred to as "element analysis" (Newton & Spurrell, 1967) and "component analysis" (Wisler, 1969), is defined as a "procedure for decomposing R^2 in multiple regression analyses into the percent of variance in the dependent variable associated with each independent variable uniquely, and the proportion of explained variance associated with the common effects of predictors" (Seibold & McPhee, 1979, p. 355). In addition, they contend that decomposing R^2 into its constituent parts is essential because:

> Advancement of theory and the useful application of research findings depend not only on establishing that a relationship exists among predictors and the criterion, but also upon determining the extent to which those independent variables, singly and in all possible combinations, share variance with the dependent variable. Only then can we fully know the relative importance of the independent variables with regard to the dependent variable in question. (p. 355)

Unlike other variance partitioning methods (e.g., stepwise regression) that distort the results by selecting variables that are not necessarily the best predictors for a particular model (Snyder, 1991; Thompson, 1995), commonality analysis considers all possible orders of entry into the model. In addition, commonality analysis is fairly safe because it does not depend on a priori knowledge to arrange the predictors. The benefit of not needing a priori knowledge is particularly important when the knowledge is fallible. According to Cooley and Lohnes (1976), "The

commonality partitioning method is neutral, and its neutrality allows the information inherent in the data about the value of organizing observations in a certain framework (that of the domain of predictors) to emerge" (p. 219).

Writing Commonality Formulas

The purpose of the present paper is to provide a brief introduction to commonality analysis as a method of variance partitioning. More detailed explanations on the derivation and calculation formulas are provided by Beaton (1973), Pedhazur (1982), or Seibold and McPhee (1979).

The unique contribution (U) of a predictor variable is defined as the portion of the variance that can be attributed to that predictor when it is entered last into the equation (Pedhazur, 1982). When there are two predictor variables the unique contribution can be expressed as:

$$U_1 = R^2_{Y.12} - R^2_{Y.2}$$

where U_1 = the unique contribution of variable 1, $R^2_{Y.12}$ = the squared multiple correlation between the dependent variable (Y) and variables 1 and 2, $R^2_{Y.2}$ = the squared multiple correlation between Y and variable 2. Similarly, the unique contribution of variable 2 is:

$$U_2 = R^2_{Y.12} - R^2_{Y.1}$$

where U_2 = the unique contribution of variable 2; $R^2_{Y.12}$ = the squared multiple correlation between Y and variables 1 and 2, $R^2_{Y.1}$ = the squared multiple correlation between Y and variable 1. Unique contributions are basically the squared semi-partial correlation between the dependent variable and the variable of interest after the effects of all other variables have been partitioned

out.

However, a commonality analysis also considers the fact that portions of the total explained variance may be common to two or more variables (Seibold & McPhee, 1979). The portion of the explained variance that is attributed to a particular group is called the common component. The common component for a two-variable model is defined as:

$$C_{12} = R^2_{y.12} - U(1) - U(2)$$

where C_{12} = the common component of variables 1 and 2. The equation can be modified with the right-hand side of the equations presented earlier for unique contributions and written as:

$$C12 = R^2_{y.12} - (R^2_{y.12} - R^2_{y.2}) - (R^2_{y.12} - R^2_{y.2})$$

$$= R^2_{y.12} - R^2_{y.12} + R^2_{y.2} - R^2_{y.12} + R^2_{y.1}$$

$$= R^2_{y.1} + R^2_{y.2} - R^2_{y.12}$$

The common component of variables 1 and 2 is called a second-order commonality. In addition, third-order common components are determined for all sets of three variables, fourth-order common components are determined for all sets of four variables, and so forth as variables are added to the model. The number of components into which the explained variance can be decomposed is equal to $2^n - 1$, where n is the number of independent variables in the regression analysis.

Therefore, the difficulty of commonality analyses increases in exponential proportion to the increases in predictor variables. For instance, in the case of a model with six predictors, R^2 can be decomposed into 64 different components. Of the 64, 58 are common components and six are unique components. One way of bypassing this problem is to arrange the variables into common

groups, then run the commonality analysis on the different groups rather than separately on individual variables. However, the problem with this method is that if the predictors are conceptually distinct, grouping them may not make any sense. Consequently, if the predictors are highly correlated, it would be extremely difficult to justify separate analyses for each variable (Pedhazur, 1982, Seibold & McPhee, Thompson, 1984). Table 1 presents the formulas for conducting commonality analyses on 2-, 3-, and 4- variable models.

To conduct a commonality analysis it is necessary that the R^2 values be computed for all possible combinations of the predictor variables (or the predictor variable groups). Rowell (1991, 1996) suggested that researchers interested in conducting a commonality analysis use the SAS (PROC SQUARE) program that will print out all possible R^2 combinations for the independent variables in the model. If SAS is unavailable, researchers can compute all necessary R^2's by individually computing each R^2 combination with SPSS. The Appendix shows all the possible R^2 combinations necessary for a three variable commonality analysis. These R^2's will be used in the heuristic example of the upcoming section. In any case, whether one uses SAS or SPSS, the procedure for calculating the necessary R^2's is fairly easy, especially if a microcomputer spreadsheet program is used.

Heuristic Example

The data set from a study conducted by Holzinger and Swineford (1939) is used here to illustrate the procedures involved in conducting a commonality analysis. The data set consists of 27 different variables obtained on 301 participants on various cognitive tests. For heuristic purposes the author arbitrarily chose scores from a General Information Verbal test (GIV) to be

the dependent variable, and scores from a Paragraph Comprehension test (PC), Word Meaning test (WM), and grade level (GL) to serve as the predictor variables.

The first step is to obtain the seven equations necessary to compute the unique and common components for a 3-variable model (see Table 1). As mentioned previously, commonality analyses can be conducted with more than three variables, the author chose a 3-variable model because it illustrates the statistical procedure without being computationally exhausting. The next step is to determine all the necessary R^2 values required for the equations in step 2 and to arrange them in tabular form. The Appendix presents all the R^2 combinations for the predictor variables used in this model.

Next all the unique and common components are determined. First, the researcher should substitute all the appropriate R^2 values into the pertaining formulas. Any spreadsheet program (e.g., Quattro Pro or Micro-Soft Excel) can be used to perform the calculations. An example of the calculations is:

$$U1(PC) = R^2_{y.123} - R^2_{y.23}$$

$$= (.5886) - (.5561)$$

$$= .0325$$

Thus, the unique contribution of variable 1, scores on the Paragraph Comprehension test, to the total explained variance is .0325, or approximately 3%. Furthermore, the common component of the Paragraph Comprehension test predictor with word meaning is:

$$C12 = -R^2_{y.3} + R^2_{y.13} + R^2_{y.23} - R^2_{y.123}$$

$$= -.0456 + .4379 + .5561 - .5886$$

$$= .3598 .$$

Therefore, the common variance accounted for by the shared contributions of the paragraph comprehension test and the word meaning test is .3598, or approximately 36%. The last step is to arrange all the unique and common components into a commonality table, such as the one presented in Table 2. Once in tabular form, the results of the unique and common components calculated can be checked by summing down each column to obtain the r^2 value when only one variable is entered into the regression equation. For instance, summing down column 3 (grade) results in a value of .0456, which is the r^2 value for the regression model when only the variable, grade level, is entered. In addition, the sum of all the unique and common components should equal to .5886, which is the R^2 value when all three predictor variables are entered into the regression model.

Discussion

An inspection of Table 2 indicates that the unique predicted variance of the Word Meaning test is approximately 15% (.1507), and its total common component with one or more of the other predictor variables is approximately 40% (.3958). In this particular model scores from the Word Meaning test account for 55% (.5465) of the variance and are considered the best predictor of performance on the General Information Verbal test. In addition, scores on the Paragraph Comprehension test uniquely accounts for 3% (.0325) of the total variance, however the total common component is approximately 40% (.3994). The other variable, grade level, contributes little to the total variance, a unique contribution of .00051 and total common component of .0405.

Although not illustrated in this heuristic example, commonality analyses can occasionally

result in some components obtaining negative values. These negative values should not be interpreted as a variable's ability to explain less than 0% of the variance (Pedhazur, 1982; Seibold, & McPhee, 1979; Thompson, 1985). Instead, the presence of a negative value is usually attributed to the presence of suppressor effects.

Commonality analysis is one method of partitioning variance in regression analyses. When there are no more than four predictor variables the analysis is fairly easy and straightforward. Commonality analysis should be very useful to educational and social science researchers when constraints restrict the number of predictors that can be used in a model. Commonality analysis is an excellent method for partitioning the variance of the dependent variable into its constituent components and for understanding the relationships of the predictors with each other and with the criterion variable.

References

Beaton, A. E. (1973). Commonality. Princeton, NJ: Educational Testing Service. (ERIC Document Reproduction Service No. ED 111 829)

Cohen, J. (1968). Multiple regression as a data-analytic system. Psychological Bulletin, 70, 426-443.

Cooley, W. W., & Lohnes, P. R. (1976). Evaluation research in education. New York: Irvington.

Elmore, P.B., & Woehlke, P. L. (1988). Statistical methods employed in American Educational Research Journal, Educational Researcher, and Review of Educational Research from 1978 to 1987. Educational Researcher, 17 (9), 19-20.

Goodwin, L. D., & Goodwin, W.L. (1985). Statistical techniques in AERJ articles, 1979-1983: The preparation of graduate students to read the educational research literature. Educational Researcher, 14 (2), 5-11.

Holzinger, K. J., & Swineford, F. (1939). A study in factor analysis: The stability of bi-factor solution (No. 48). Chicago, IL: University of Chicago. (Data on pp. 81-91).

Murthy, K. (1994, April). Commonality analysis for the regression case. Paper presented at the annual meeting of the American Educational Research Association, New Orleans. (ERIC Document Reproduction Service No. ED 367 694)

Neter, J., Kutner, M. H., Nachtsheim, C. J., & Wasserman, W. (1996). Linear regression with one predictor variable. In Applied linear regression (3rd ed. pp. 3-43). Chicago: Irwin.

Newton, R. G., & Spurrell, D. J. (1967). Examples of the use of elements for classifying regression analysis. Applied Statistics, 16, 165-172.

Pedhazur, E. J. (1982). Multiple regression in behavioral research: Explanation and prediction (2nd ed.). New York: Holt, Rinehart, and Winston.

Rowell, R. K. (1991, January). Partitioning predicted variance into constituent parts: How to conduct commonality analysis. Paper presented at the annual meeting of the Southwest Educational Research Association, San Antonio. (ERIC Document Reproduction No. ED 328 589)

Rowell, R. K. (1996). Partitioning predicted variance into constituent parts: How to conduct commonality analysis. In B. Thompson (Ed.), Advances in social science methodology (Vol. 4, pp. 33-44). Greenwich, CT: JAI Press.

Seibold, D. R., & McPhee, R. D. (1979). Commonality analysis: A method for decomposing explained variance in multiple regression analyses. Human Communication Research, 5, 355-366.

Snyder, P. (1991). Three reasons why stepwise regression methods should not be used by researchers. In B. Thompson (Ed.), Advances in educational research: Substantive findings, methodological developments. (Vol. 1, pp. 99-106). Greenwich, CT: JAI Press.

Thompson, B. (1984, November). Coding and commonality analysis: Non-ANOVA methods for analyzing data from experiments. Paper presented at the annual meeting of the Mid-South Educational Research Association, New Orleans. (ERIC Document Reproduction No. ED 257 881)

Thompson, B. (1985). Alternative methods for analyzing data from education experiments. Journal of Experimental Education, 54, 50-55.

Thompson, B. (1992, April). Interpreting regression results: beta weights and structure coefficients are both important. Paper presented at the annual meeting of the American Educational Research Association, San Francisco. (ERIC Document Reproduction No. ED 344 897)

Thompson, B. (1994, April). Common methodology mistakes in dissertations, revisited. Paper presented at the annual meeting of the American Educational Research Association, New Orleans. (ERIC Document Reproduction No. ED 368 771)

Thompson, B. (1995). Stepwise regression and stepwise discriminant analysis need not apply here: A guideline editorial. Educational and Psychological Measurement, 55, 525-534.

Thompson, B., & Borrello, G.M. (1985). The importance of structure coefficients in regression research. Educational and Psychological Measurement, 45, 203-209.

Willson, V. L. (1980). Research techniques in AERJ articles: 1969 to 1978. Educational Researcher, 9 (6), 5-10.

Wisler, C. E. (1969). On partitioning the explained variation in a regression analysis. Washington D.C.: Office of Education/ U.S. Department of Health, Education, and Welfare. (ERIC Document Reproduction No. ED 174 408)

Table 1

Formulas for Unique and Commonality Components of Variance

Two Independent Variables

$U(1) = R^2_{y.12} - R^2_{y.2}$
$U(2) = R^2_{y.12} - R^2_{y.1}$
$C(12) = R^2_{y.2} + R^2_{y.1} - R^2_{y.12}$

Three Independent Variables

$U(1) = R^2_{y.123} - R^2_{y.23}$
$U(2) = R^2_{y.123} - R^2_{y.13}$
$U(3) = R^2_{y.123} - R^2_{y.12}$
$C(12) = - R^2_{y.3} + R^2_{y.13} + R^2_{y.23} - R^2_{y.123}$
$C(13) = - R^2_{y.2} + R^2_{y.12} + R^2_{y.23} - R^2_{y.123}$
$C(23) = - R^2_{y.1} + R^2_{y.12} + R^2_{y.13} - R^2_{y.123}$
$C(123) = R^2_{y.1} + R^2_{y.2} + R^2_{y.3} - R^2_{y.12} - R^2_{y.13} - R^2_{y.23} + R^2_{y.123}$

Four Independent Variables

$U(1) = R^2_{y.1234} - R^2_{y.234}$
$U(2) = R^2_{y.1234} - R^2_{y.134}$
$U(3) = R^2_{y.1234} - R^2_{y.124}$
$U(1) = R^2_{y.1234} - R^2_{y.123}$
$C(12) = - R^2_{y.34} + R^2_{y.134} + R^2_{y.234} - R^2_{y.1234}$
$C(13) = - R^2_{y.24} + R^2_{y.124} + R^2_{y.234} - R^2_{y.1234}$
$C(14) = - R^2_{y.23} + R^2_{y.123} + R^2_{y.234} - R^2_{y.1234}$
$C(23) = - R^2_{y.14} + R^2_{y.124} + R^2_{y.134} - R^2_{y.1234}$
$C(24) = - R^2_{y.13} + R^2_{y.123} + R^2_{y.134} - R^2_{y.1234}$
$C(34) = - R^2_{y.12} + R^2_{y.123} + R^2_{y.124} - R^2_{y.1234}$
$C(123) = - R^2_{y.4} + R^2_{y.14} + R^2_{y.24} + R^2_{y.34} - R^2_{y.124} - R^2_{y.134} - R^2_{y.234} + R^2_{y.1234}$
$C(124) = - R^2_{y.3} + R^2_{y.13} + R^2_{y.23} + R^2_{y.34} - R^2_{y.123} - R^2_{y.134} - R^2_{y.234} + R^2_{y.1234}$
$C(134) = - R^2_{y.2} + R^2_{y.12} + R^2_{y.23} + R^2_{y.24} - R^2_{y.123} - R^2_{y.124} - R^2_{y.234} + R^2_{y.1234}$
$C(234) = - R^2_{y.1} + R^2_{y.12} + R^2_{y.13} - R^2_{y.14} - R^2_{y.123} - R^2_{y.124} - R^2_{y.134} + R^2_{y.1234}$
$C(1234) = R^2_{y.1} + R^2_{y.2} + R^2_{y.3} + R^2_{y.4} - R^2_{y.12} - R^2_{y.13} - R^2_{y.14} - R^2_{y.23} - R^2_{y.24} - R^2_{y.34} + R^2_{y.123} + R^2_{y.124} + R^2_{y.134} + R^2_{y.234} - R^2_{y.1234}$

Note. The difficulty of interpretation increases in proportion to the increases in predictor variables.

Table 2

Commonality Analysis Summary Table

Component	1 Paragraph Com	2 Word Meaning	3 Grade Level
U(1)	.0325		
U(2)		.1507	
U(3)			.0051
C(12)	.3598	.3598	
C(13)	.0045		.0045
C(23)		.0009	.0009
C(123)	.0351	.0351	.0351
Total	.4319	.5465	.0456
U	.0325	.1507	.0051
C	.3994	.3958	.0405

Note. The sum of the columns equals the R^2 of that particular predictor and the sum of all the unique and common components equals the multiple R^2 of the regression equation.

Appendix

R-Squares of Paragraph Comprehension Test, Word Meaning Test, and Grade to General Information Verbal Test

Number of Procedures in Model	R-square	Variables in Model
1	.4319	1 Paragraph Comprehension test
	.5468	2 Word Meaning test
	.0456	3 grade level
2	.5835	12
	.4379	13
	.5561	23
3	.5886	123